Two Computational Primitives for Algorithmic Self-Assembly: Copying and Counting

Supporting Information

Robert D. Barish,[*] Paul W. K. Rothemund,[†,‡] Erik Winfree[*,†]

Contribution from the Departments of Computer Science and Computation and Neural Systems,
California Institute of Technology, Pasadena, CA 91125

E-mail:winfree@dna.caltech.edu

1 Methods

Estimated bound on effective error rates:

Given an assembly of tiles on a scaffold of N repeat units in which each tile has a mismatch with its lower or right neighbor with independent probability $\epsilon = 1 - p$, we would like to bound the probability that the assembly contains a particular pattern of M tiles that has been observed. In particular, we are interested in the probability of seeing *any* pattern with fewer than m mismatched tiles. At a particular location along the scaffold, the probability that the relevant M tiles have fewer than m mismatches is

$$Pr(less\ than\ m\ errors\ at\ site\ i) = \sum_{n=0}^{m-1} \binom{M}{n} p^{M-n} \epsilon^n.$$

The pattern could appear at up to N positions along the scaffold, and although the probabilities of observing the pattern within overlapping regions are not independent, we can use a union-sum bound:

$$Pr(less\ than\ m\ errors\ at\ some\ site)$$
$$\leq \sum_{i=1}^{N} Pr(less\ than\ m\ errors\ at\ site\ i).$$

For $\epsilon = 0.15$ and $N = 2400$ and $M = 126$ and $m = 4$, the probability of finding at least one low-error pattern is less than

$$2400 \sum_{n=0}^{3} \binom{126}{n} 0.85^{126-n} 0.15^n \approx 0.0063.$$

UV photometry formation and melting experiments:

An AVIV Model 14NT UV/VIS spectrophotometer was used to measure thermal formation and melting profiles at 260 nm. To avoid formation of air bubbles at elevated temperature, samples were vacuum-degassed before use. To measure thermal formation profiles, samples were heated to 90°C and cooled to 20°C using 0.1°C temperature steps with a 0.1°C temperature deadband and an equilibration time of 0.25 minutes. After 20°C was reached, the melting profile was measured while heating from 20°C to 90°C using the same parameters.

Generation of the Long Scaffold Strand:

To create the long scaffold strand, assembly PCR was used. The desired repetitive 62-mer sequence was conceptually divided into two subsequences, a 30-mer and a 32-mer (Figure S1, top). Primers for assembly PCR consisted of these two oligonucleotides (SCA-Splint1 and SCA-Splint2) as well as two complementary overlapping oligonucleotides (SCA-Comp1 and SCA-Comp2). In each cycle of assembly PCR, complementary primers or products from the previous cycle bind to each other, allowing polymerase to extend strands wherever a $5'$ overhang can serve as a template (Figure S1, middle). After many cycles, this leads to roughly exponential elongation and results in long, mostly double-stranded product.

The hairpin subsequence within the 62-mer repeat sequence poses a potential problem: a primer or product that forms a hairpin with a duplex stem containing a terminal base pair might serve to prime polymerization, which would result in an incorrect sequence being produced. To try to avoid this problem, the position of the primer subsequences within the 62-mer repeat sequence was chosen such that the hairpin subsequence of the scaffold strand (Figure S2, bottom) is divided between the SCA-Splint1 and SCA-Splint2 primers. On SCA-Comp1, the 3-nt and 7-nt regions beyond the stem of the hairpin, at the $3'$ and $5'$ ends of the oligonucleotide respectively, should prevent incorrect initialization of polymerases and serve as 'toe-hold' region allowing for complementary primers or products to bind. Nonetheless, the second primer-primer complex shown in Figure S1 for the first cycle of assembly PCR may form much less frequently than the other primer-primer complex. (If it did not form at all, assembly PCR would not continue.)

Because double-stranded products are ineffective for serving as a scaffold for tile assembly, we needed a way to enrich for the scaffold strand sequence and not its complement. We chose to do this using asymmetric PCR in the last (4^{th}) stage of assembly PCR, wherein only the nucleotides dATP, dTTP, and dCTP are provided, in addition to new primers. Thus, by designing the the scaffold strand sequence to use only the adenine, thymine, and cytosine bases (and consequently, the hairpin domains in the scaffold are entirely A and T), only the scaffold strand sequence is extended during the last stage of assembly PCR. This stage in principle results in linear rather than exponential growth of the product material, and therefore we expect a significant fraction of the final product to still be double-stranded.

The assembly PCR procedure consists of three exponential amplification stages, followed by one linear extension stage. In the first stage, 1.0 pmole of total primers and splints (1/N pmole of each) is added to a 19 μL reaction mixture consisting of 1.6 μL dNTPs (10 mM dNTPs, 2.5 mM each), 1 μL Magnesium Acetate (25 mM), 6 μL GeneAMP XL PCR Buffer (3.3 X buffer), 10 μL water. After the mixture is incubated at 40C for three minutes, 0.4 μL of GeneAMP XL polymerase is added to bring the total volume of to 20 μL. The stage 1 temperature program consists of 40 cycles of a 94°C step for 15 seconds, a 40°C step for 30 seconds, and a 72°C for 10 seconds with an additional one second per cycle. In stage 2, no new primers or splints are added and the total volume is brought to 60 μL with the addition of 40 μL of fresh reaction mixture and polymerase. The temperature program for stage 2 consists of 25 cy-

[*] California Institute of Technology
[†] Department of Computer Science, California Institute of Technology
[‡] Computation and Neural Systems, California Institute of Technology

cles of a 94°C step for 15 seconds, a 40 °C step for 30 seconds, and a 72°C step for 45 seconds with an additional one second per cycle. For the final exponential amplification stage, the 60 μL reaction mixture is split into three 20 μL volumes, to which one again adds 40 μL of fresh reaction mixture and polymerase. The stage 3 temperature program consists of 20 cycles with a step at 94°C for 15 seconds, a step at 40°C for 30 seconds, and a step at 72°C for 70 seconds with one addition second per cycle. For the final PCR linear amplification stage, 5 μL of stage 3 product is added to 55 μL of reaction mixture with only dATP, dTTP, and dCTP nucleotides. This prevents the extension of the splint strands and creates an excess of long single stranded scaffold. The temperature program used for stage 4 consists of 60 cycles of the stage 3 temperature program.

For all PCR reactions an Stratagene Mx4000 quantitative PCR (qPCR) machine was used. Fluorescence of Sybr Green I dye, at a concentration of 1:100000, was used to monitor the progress of the reactions. Sybr Green I fluorescence increases strongly upon binding double-stranded DNA and increases somewhat less strongly (10-fold less) upon binding single-stranded DNA. Thus Sybr Green I fluorescence is used widely as a surrogate for the amount of double-stranded product in quantitative PCR reactions and a characteristic sigmoidal curve composed of an exponential growth phase followed by saturation is typically observed. Here, only in PCR stage I did fluorescence exhibit the sigmoidal growth curve typical of normal PCR reactions. For stages 2, 3, and 4 fluorescence saturated quickly without an initial exponential growth phase. We do not attempt to explain these curves but merely emphasize that normal fluorescence curves should not be expected for these stages.

A phenol/chloroform extraction is performed to remove proteins from the PCR reaction mixture. Samples are diluted to 100 μL and two volumes of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma P2069) are added. Instead of vortexing, the samples are shaken gently by hand for two minutes to prevent shearing. (We have not demonstrated that such treatment actually decreases shearing.) After removing the phenol/chloroform phase, two volumes of pure chloroform are then added to the aqueous phase and the same procedure is performed.

Following phenol/chloroform extraction, an ethanol precipitation is performed to remove salts and the Sybr-Green I dye from the sample. 2/3X volumes of 7.5 M ammonium acetate is added to the each sample followed by 2X the current volume of 200 proof 'molecular biology grade' ethanol. After staying at -72°C for two hours, samples are centrifuged at 12000 rcf at 3°C. Each sample is then decanted and 200 μL of cold 70% ethanol is added. After gently shaking by hand for two minutes, they are centrifuged for an additional 15 minutes at 3°C 12000 rcf. Following the second centrifugation, each sample is decanted, resuspended, and quantitated using the Eppendorf BioPhotometer. Stock solutions are prepared at 30 nM as described in the main text of the paper.

Correspondence should be addressed to Erik Winfree (winfree@caltech.edu).

References

(1) Rothemund, P. W. K.; Ekani-Nkodo, A.; Papadakis, N.; Kumar, A.; Fygenson, D. K.; Winfree, E. *Journal of the American Chemical Society* **2004,** *126,* 16344–16352.

(2) Puglisi, J. D.; Tinoco, Jr., I. . In *Methods in Enzymology*, Vol. 180; Dahlberg, J. E.; Abelson, J. N., Eds.; Academic Press: San Diego, 1989.

To create the long SCA scaffold strand, the 62-mer repeat is divided into a 30 and 32-mer splint sequence as shown below:

SCA Splints:

SCA-Splint1 (30-mer): 5' - TATAATATCCCATACCTATCACTAACCCCT - 3'

SCA-Splint2 (32-mer): 5' - CATCCTCTCTCCCACCCTACATATTATATTTT - 3'

```
5' - ...TATAATATCCCATACCTATCACTAACCCCTCATCCTCTCTCCCACCCTACATATTATATTTTTATAATATCCCATACCTATCACTAACCCCT... - 3'
      <-------------------------------------------------------------->
                              62-mer repeat
```

In addition,two complementary splints with 30 and 32-base complementary are used for assembly PCR in exponential amplification stages 1-3:

SCA Complementary Splints:

SCA-Comp1 (30-mer): 3' - ATGTATAATATAAAAATATTATAGGGTATG - 5'

SCA-Comp2 (32-mer): 3' - GATAGTGATTGCGGAGTAGGAGAGAGGGTGGG- 5'

1st Cycle of Assembly PCR:

```
                    5' - TATAATATCCCATACCTATCACTAACCCCT - 3' -----> Extension by Polymerase                    SCA
Extension by Polymerase <----- 3' - GATAGTGATTGCGGAGTAGGAGAGAGGGTGGG- 5'                                     SCA Comp

                    5' - CATCCTCTCTCCCACCCTACATATTATATTTT - 3' -----> Extension by Polymerase                  SCA
 Extension by Polymerase <----- 3' - ATGTATAATATAAAAATATTATAGGGTATG - 5'                                     SCA Comp
```

2nd Cycle of Assembly PCR:

```
                    5' - TATAATATCCCATACCTATCACTAACCCCTCATCCTCTCTCCCACCC - 3' -----> Extension by Polymerase    SCA
                Extension by Polymerase <----- 3' - GTAGGAGAGAGGGTGGGATGTATAATATAAAAATATTATAGGGTATG - 5'     SCA Comp

                    5' - CATCCTCTCTCCCACCCTACATATTATATTTTTATAATATCCCATAC - 3' -----> Extension by Polymerase    SCA
                  Extension by Polymerase <----- 3' - ATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGG - 5'   SCA Comp
```

Nth Cycle of Assembly PCR:

```
5' - ...TATAATATCCCATACCTATCACTAACCCCTCATCCTCTCTCCCACCCTACATATTATATTTTTATAATATCCCATACCTATCACTAACCCCTCATCCTCTCTCCCACCC... - 3'   SCA
3' - ...ATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGGATGTATAATATAAAAATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGG... - 3'   SCA Comp
```

In PCR Stage 4, the linear amplification stage, no dGTP is added to reaction mixture:

```
   5' - TATAATATCCCATACCTATCACTAACCCCT - 3' -----> Extension by Polymerase                                    SCA
3' - ...ATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGGATGTATAATATAAAAATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGG... - 3'   SCA Comp

   5' - CATCCTCTCTCCCACCCTACATATTATATTTT - 3' -----> Extension by Polymerase
3' - ...GTAGGAGAGAGGGTGGGATGTATAATATAAAAATATTATAGGGTATGGATAGTGATTGCGGAGTAGGAGAGAGGGTGGGATGTATAATATAAAAATATTATAGGGTATG... - 3'   SCA Comp
```

Figure S1. Schematic for how the repetitive SCA scaffold strand is created using assembly PCR. The 62-mer repeating sequence of the scaffold and the four primer oligonucleotides are shown at the top. Regions underlined in black on SCA-Splint1, SCA-Splint2, and SCA-Comp1 indicate the stem of the hairpin region, while the pink underline indicates the poly-A (SCA-Comp1 sequence) and poly-T loops (SCA-Splint2). In PCR stages 1-3, shown in the middle, exponential elongation of primers takes place using all four dNTP nucleotides, so that both sense and antisense sequences are synthesized. Stage 4, shown at the bottom, uses only dATP, dTTP, and dCTP nucleotides and new primers, allowing only the linear extension of the scaffold strand and producing primary the desired single-stranded molecule.

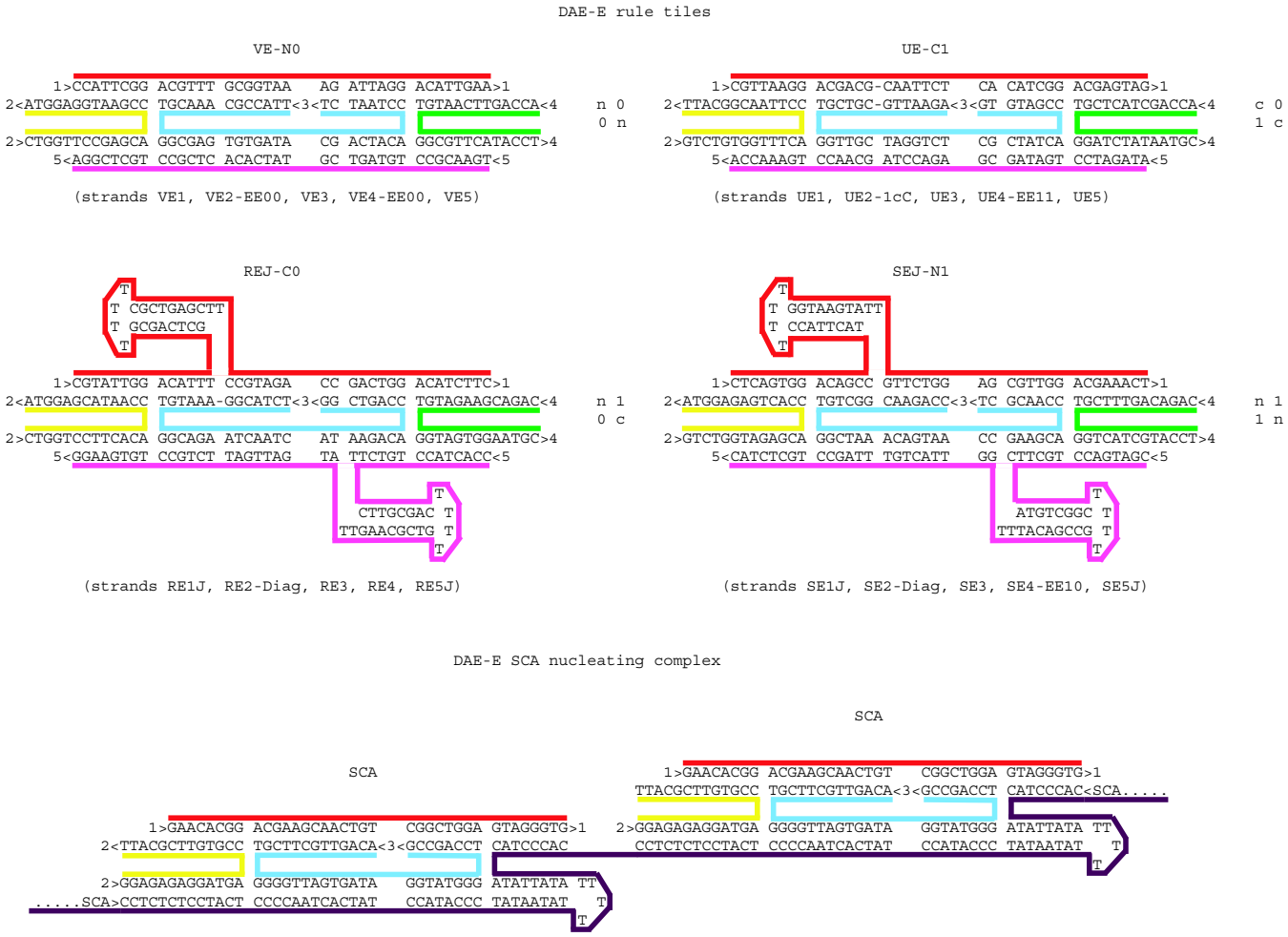

Figure S2. Schematics showing the sequences for the four DAE-E double crossover rule molecules used in this paper, the repetitive scaffold strand, and the scaffold tiles. The component strands for each DX is listed underneath the tile. Strands with the same names as those given in Ref.[1] have the same sequences as in that work; additionally, strands here containing a "J" in their name correspond to strands there with the suffix "-h14". In fact, tile VE-N0 is identical to tile VE-00 of Ref.,[1] tile SEJ-N1 is identical to tile SEs(1,5:h14), and tile REJ-C0 differs from tile REp(1,5:h14) by exactly one sticky-end only. Within each tile, strands are numbered 1 through 5 and are labeled at their 5′ and 3′ ends; wedges ("¿" or "¡") pointing toward the first nucleotide indicate 5′ ends, and wedges pointing away from the last nucleotide indicate 3′ ends. Two repeat units of the repetitive scaffold strand (SCA) are shown, along with the strands 1, 2, and 3 that compose the scaffold tile (also called SCA); our assembly PCR protocol appears to generate scaffold strands containing up to roughly 40 repeat units.

DAE-E system strands:

Rule tile strands.

VE1	(37-mer,	377840 /M/cm @ 260nm)	:	CCATTCGGACGTTTGCGGTAAAGATTAGGACATTGAA
VE2_EE00	(26-mer,	260540 /M/cm @ 260nm)	:	CTGGTTCCGAGCACCGAATGGAGGTA
VE3	(42-mer,	412740 /M/cm @ 260nm)	:	TTACCGCAAACGTGGCGAGTGTGATACGACTACACCTAATCT
VE4_EE00	(26-mer,	249800 /M/cm @ 260nm)	:	ACCAGTTCAATGTGGCGTTCATACCT
VE5	(37-mer,	348140 /M/cm @ 260nm)	:	TGAACGCCTGTAGTCGTATCACACTCGCCTGCTCGGA
UE1	(37-mer,	374540 /M/cm @ 260nm)	:	CGTTAAGGACGACGCAATTCTCACATCGGACGAGTAG
UE2_1cC	(26-mer,	243160 /M/cm @ 260nm)	:	GTCTGTGGTTTCACCTTAACGGCATT
UE3	(42-mer,	404820 /M/cm @ 260nm)	:	AGAATTGCGTCGTGGTTGCTAGGTCTCGCTATCACCGATGTG
UE4_EE11	(26-mer,	253840 /M/cm @ 260nm)	:	ACCAGCTACTCGTGGATCTATAATGC
UE5	(37-mer,	378680 /M/cm @ 260nm)	:	ATAGATCCTGATAGCGAGACCTAGCAACCTGAAACCA
RE1J	(59-mer,	553620 /M/cm @ 260nm)	:	CGTATTGGACATTTGCTCAGCGTTTTCGCTGAGCTTCCGTAGACCGACTGGACATCTTC
RE2_Diag	(26-mer,	253800 /M/cm @ 260nm)	:	CTGGTCCTTCACACCAATACGAGGTA
RE3	(42-mer,	430880 /M/cm @ 260nm)	:	TCTACGGAAATGTGGCAGAATCAATCATAAGACACCAGTCGG
RE4	(26-mer,	273000 /M/cm @ 260nm)	:	CAGACGAAGATGTGGTAGTGGAATGC
RE5J	(59-mer,	549780 /M/cm @ 260nm)	:	CCACTACCTGTCTTCTTGCGACTTTTGTCGCAAGTTATGATTGATTCTGCCTGTGAAGG
SE1J	(59-mer,	572120 /M/cm @ 260nm)	:	CTCAGTGGACAGCCTACTTACCTTTTGGTAAGTATTGTTCTGGAGCGTTGGACGAAACT
SE2-Diag	(26-mer,	267700 /M/cm @ 260nm)	:	GTCTGGTAGAGCACCACTGAGAGGTA
SE3	(42-mer,	415380 /M/cm @ 260nm)	:	CCAGAACGGCTGTGGCTAAACAGTAACCGAAGCACCAACGCT
SE4_EE10	(26-mer,	249220 /M/cm @ 260nm)	:	CAGACAGTTTCGTGGTCATCGTACCT
SE5J	(59-mer,	539060 /M/cm @ 260nm)	:	CGATGACCTGCTTCATGTCGGCTTTTGCCGACATTTGGTTACTGTTTAGCCTGCTCTAC

Input tile strands for use with SCA nucleating strand.

SCA1	(37-mer,	379280 /M/cm @ 260nm)	:	GAACACGGACGAAGCAACTGTCGGCTGGAGTAGGGTG
SCA2	(26-mer,	266040 /M/cm @ 260nm)	:	GGAGAGAGGATGACCGTGTTCGCATT
SCA3	(42-mer,	417300 /M/cm @ 260nm)	:	AGTTGCTTCGTGGGGTTAGTGATAGGTATGGGTCCAGCCGAC

Splint strands for making SCA nucleating strand with assembly PCR.

SCA_Splint1	(37-mer,	287760 /M/cm @ 260nm)	:	TATAATATCCCATACCTATCACTAACCCCT
SCA_Splint2	(37-mer,	290320 /M/cm @ 260nm)	:	CATCCTCTCTCCCACCCTACATATTATATTTT
SCA_Comp1	(37-mer,	332400 /M/cm @ 260nm)	:	GTATGGGATATTATAAAAATATAATATGTA
SCA_Comp2	(37-mer,	351460 /M/cm @ 260nm)	:	GGGTGGGAGAGAGGATGAGGGGTTAGTGATAG

Figure S3. Sequences used in the binary counter experiments. Extinction coefficients are computed using the nearest-neighbor model of Ref.[2]

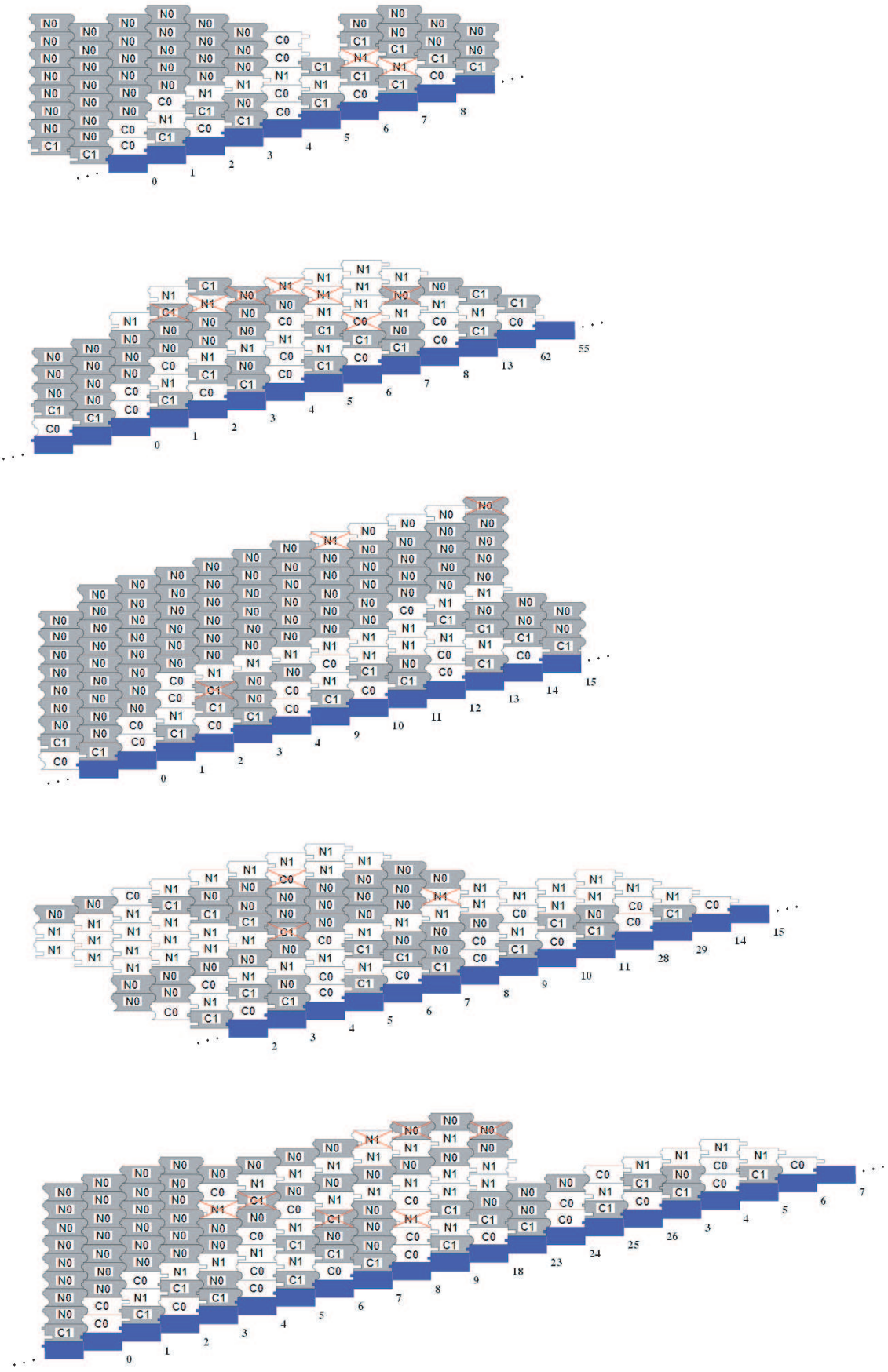

Figure S4. Detailed interpretation of the AFM images in Figure 3a. The interpreted fragments contain respectively 70, 68, 126, 102, and 120 tiles.